

ПРОФЕСІЙНА ОСВІТА

UDC 811.111:371.3 (045)

Olena P. Petrashchuk, Doctor of Pedagogy, Prof.**USEFUL ENGLISH LANGUAGE TEST IN AVIATION**National Aviation University
E-mail: aamm@nau.edu.ua

Розглянуто проблему придатності лінгво-дидактичного тесту загального володіння мовою (Proficiency Test) як інструменту для оцінювання рівня сформованості англомовної складової професійної компетентності фахівця авіаційної галузі. Описано якісні характеристики та специфікації цього тесту. Запропоновано шляхи визначення ступеня придатності тесту з англійської мови для його застосування в авіації.

The article deals with the issue of test usefulness, namely usefulness of a proficiency language test as a tool to assess English language component of a professional competency required for aviation personnel. The test quality characteristics and specifications are described. The ways to evaluate the test usefulness with the purpose to further use in aviation are presented.

Рассмотрена проблема приемлемости лингво-дидактического теста общего владения языком (Proficiency Test) как инструмента для оценки уровня сформированности англоязычной составной профессиональной компетентности специалиста авиационной отрасли. Описаны качественные характеристики и спецификации такого теста. Предложены пути определения степени приемлемости теста по английскому языку для его применения в авиации.

Statement of Purpose

Standards for selection and/or development of language tests should be identified when a high stake proficiency test is recommended as a tool to assess language competence in a specific field. In my view the English language use in aviation is the case when a useful language test is an appropriate instrument to measure a language competence of aviation personnel.

It should be underlined that the language competence is considered as an integrant component of the entire professional competency. Therefore quality of the test is a key point to obtain reliable and accurate results of the measurement. The results can be used for credible judgements which have a great impact in high stake testing.

To get usefulness of a test six qualities should be provided. The test development should be conducted in accordance with its specifications which are outlined prior to the

development procedure. In order to check the degree of the test usefulness a specially designed checklist can be applied to ensure the expected quality of the test.

Review of Research Results

The qualities of a language test have been discussing for long time though it is clear that there can be no good or bad test. It is a matter of test purpose with which the test has been developed.

The same can be stated for language tests in aviation: there is no “perfect” aviation English test. The purpose of current English language testing in aviation is licensure for pilots and controllers.

The international requirement for civil pilots and controllers to demonstrate, as of March 2008 (March 2011 – final deadline), a minimum level of proficiency in the languages used for aviation radiotelephony communications - and to do so repeatedly throughout their careers -

has established an unprecedented need worldwide for language testing procedures and systems that can elicit job-specific (radiotelephony communications) language performances which can be related to all the categories defined in the ICAO language proficiency rating scale [1].

I agree with Dr. J.Mell that the current situation with regard to fulfilling this need internationally is widely recognised as being unsatisfactory, resulting in local searches by those invested with this responsibility for solutions ranging from “off the peg” general tests to development of special purpose tests. Meanwhile, the signs are that the language testing industry itself is gearing up to be able to provide solutions, and, in the absence of any independent international oversight, the aviation community must prepare itself to be able to assess the suitability of what is on offer [2].

For example, in aviation a high-stake proficiency test is required to be used for assessment of English language proficiency of pilots and air traffic controllers against ICAO scale. It is obvious that such a test should be of high quality.

According to J.Mell the main characteristics of such a test are as follows [2]:

- 1) items are chosen to resemble real-world tasks;
- 2) overall scores are holistic: YES/NO;
- 3) stakes: very high;
- 4) a “blunt” instrument but must be highly robust:
 - proven validity;
 - proven reliability;
 - security

According to L. Bachman and A. Palmer when we design (or select) a language test we need to consider the characteristics of the language use situation and tasks and of language users and test takers. These two sets of characteristics that affect both language use and language test performance are of central interest [3].

In order to provide high quality measurement of language competence the characteristics of test language performance should correspond with characteristics of language use in real life. So, a language test should be useful for a particular testing situation.

Purpose of the work

It is obvious that language test results in aviation have a great impact in both professional and social areas. Therefore the test usefulness seems to be an important consideration.

This is especially important due to new ICAO language requirements which in addition to standard radiotelephony phraseology, stands for ability to use plain English in aviation context.

Regarding safety of flights it is clear that the more language test performance measurement is reliable and accurate the more exactly a level of language use in real life is predicted. In other words, for language assessment purposes in aviation the correspondence between language test performance and non-test language use is a key issue.

L. Bachman gives a clear and explicit definition of the qualities of test usefulness: “Test usefulness, consisting of several qualities (reliability, construct validity, authenticity, interactiveness, impact, practicality), is an overriding consideration for quality control throughout the process of designing, developing, and using a particular language test” [3, p. 9].

Correspondence between language use and language test performance of aviation personnel will comprise the following sets of characteristics:

1. Characteristics of the language use task and situation and characteristics of the test task and situation. It is clear that the tasks and the situations will be job related taking into account specificity of pilots’ and controllers’ job places.

2. Language use in real life and language test performance.

3. Characteristics of the language user (a pilot/controller) and characteristics of the test taker. These characteristics will cover topical knowledge outlined in aviation lexical domains, affective schemata and language ability presented in the ICAO scale descriptors.

The test usefulness provides a kind of metric by which a test can be evaluated. That’s why a test usefulness is regarded as the essential basis

for quality control throughout of entire test development process. At the same time usefulness cannot be evaluated in the abstract for all tests [3].

Aforementioned test usefulness is based on six test qualities – reliability, construct validity, authenticity, interactiveness, impact, practicality. It should be mentioned that in ICAO documents only three of the above qualities are recommended for the test, namely reliability, validity and practicality [4; 5]. Though the test usefulness might not be provided by the three qualities, in this article those three will be considered.

For aviation purposes one cannot say what the appropriate balance among the different test qualities should be or what are minimum acceptable levels. Evaluating the overall usefulness of a given test is essentially subjective due to the judgements made by a test designer [3].

However with a reference to ICAO recommendations [4], it can be presumed that a test designer may wish to design the test and test tasks so as to achieve the highest possible levels of reliability and validity because of a high-stake status of the language test for aviation.

It should be underlined that traditionally researchers reported about the test qualities being in conflict. I agree with L. Bachman and A. Palmer that the testers should recognize the test qualities complementarity rather than their tension [3].

Regarding the test qualities that determine the test usefulness it is essential to take a systemic view, considering tests as part of a larger societal and educational context [3].

Reliability is defined as consistency of measurement and can be considered to be a function of the consistency of scores from one set of tests and test tasks to another as well as a function of consistencies across different sets of test tasks characteristics [3, p. 20].

It is clear that reliability is an essential quality of test scores, for unless test scores are relatively consistent, they cannot provide us with any information at all about ability we

want to measure, e.g., pilots'/controllers' ability to use English language at operational level. Taking into account that this ability is measured by oral proficiency interview¹ reliability of raters will also have great influence on the test reliability.

If some raters rate more severely than others, then the ratings of different raters are not consistent, and the scores obtained could not be considered to be reliable [3].

Therefore in aviation context in addition to test tasks reliability special training of raters on a regular base should be provided to establish and further maintain their intra- and inter-reliability.

Validity of a test deals with meaningfulness and appropriateness of the interpretations that one makes on the basis of test scores. To what extent these interpretations can be justified to be valid indicators of test takers' language ability is a crucial question [3, p.21].

Appropriate justification of a particular score interpretation is based on the evidence that the test score reflects the areas of language ability which is measured. For aviation purposes a construct can be considered the specific definition of an ability that provides the basis for a given test or test task and for interpreting scores derived from this task. Therefore the term construct validity can be used to refer to the extent to which a given test score could be interpreted as an indicator of the ability (construct) which is measured. If the construct validity of a score interpretation is considered then both the construct definition and the characteristics of the test task should be considered. In this respect there is a need to determine the extent to which the test task corresponds to tasks in the target language use domain.

It would appear, on the basis that pilots and controllers participate in the same communicative discourse that great amounts of time and money could be saved if the same test could be used for both populations targeted by

¹ Here I refer to the test format developed by me for AEROLINGUA company.

ICAO language proficiency requirements. For language testing purposes though, just as for evaluation of other respective professional competencies, this would pose a number of problems.

Pilots and controllers play different, albeit complementary, interactive roles in radiotelephony communications. If a test taker is to demonstrate interactive ability in a job-related scenario, this role will govern the design of the testing format, which will have to be different for each population.

Additionally, the passive and productive language repertoires of pilots and controllers are complementary. Stated simply, this means that what controllers need to be able to understand, pilots need to be able to say, and that what pilots need to be able to understand, controllers need to be able to say [2].

This distinction may result in differing testing objectives for each population in terms of the lexical and functional content of listening and speaking tasks thus identifying different construct validity.

In order to have construct validity the language test for pilots/controllers should be designed with a reference to target language proficiency according to ICAO analytic scale descriptors. Therefore the test (test tasks) is to:

- provide a representative range of intelligible international accents as input for comprehension;
- provide a professionally relevant format for candidates to display comprehension;
- elicit an adequate continuous speech sample to test fluency/pronunciation;
- provide a voice-only setting for “diadic” (2-person) interactions;
- provide examples of routine and unexpected events in a work-related context;
- to use basic grammatical structures creatively;
- allow the candidate to demonstrate ability to paraphrase;
- allow the candidate to change between rehearsed/formulaic speech and spontaneous interaction;

– simulate unexpected events to create opportunities for misunderstanding [2].

Practicality pertains primarily to the ways in which the test is implemented in a given situation [3, p.39]. In other words, practicality may be defined as the balance between the resources that will be required in the design, development and use of the test and the resources that will be available for these activities. Several types of resources can be identified: human resources, material resources and time. Thus, practicality can only be determined for a specific testing situation, for example in aviation.

Consideration of means of test delivery will certainly lead to the evaluation of appropriate technologies.

The voice-only character of radiotelephony communications means that telephones, 2-way radios, language laboratories or training simulators would all be preferred environments for mediating interactions.

Delivery of test input (text, sound, or graphics) is clearly simplified by computerisation. Similarly, candidate spoken performances are more rapidly accessed for rating if they are stored as digital sound files rather than recorded on analogue tape.

Finally, the possibility for speech recognition technologies to automate the rating of spoken performances is a tantalising prospect, but their acceptability would depend on prior standardisation of the parameters of “intelligibility” and they would need to demonstrate their capacity to filter all the possible speech variations for the region in which such testing technology is intended to be used.

In any case, the choice of delivery technologies will usually have to be aligned to the lowest level of equipment available to the target users. Additionally, current proven computer technology is unable to simulate the natural features of interaction. It is difficult to see how an ICAO-compliant test could exclude a human-human encounter allowing opportunities to evaluate the ability to initiate

and maintain exchanges and to deal with misunderstandings. It is significant that all specific tests developed so far have maintained this component [2].

In order to make language testing for aviation more or less useful test designers could have benefits from the Bachman's and Palmer's philosophy of language testing [3, p. 13] consisting of the following six principles:

1. Relate language testing to language teaching and language use.

In aviation the principle can be applied to the ICAO recommendations for new language requirements implementation [4; 5; 6] which is based on complex actions including training, testing and operational language level maintenance.

2. Design your tests so as to encourage and enable tests takers to perform at their highest level of ability.

For the aviation licensure purposes it may mean the need for a pilot/controller to demonstrate his/her best language performance so that an appropriate ICAO proficiency level is conformed.

3. Build considerations of fairness into test design.

In aviation a language related human error may affect flight safety. In this respect the pilot's/controller's language proficiency assessment is a concern of fairness to satisfy the ICAO proficiency in plain English requirements.

4. Humanize the testing process: seek ways in which to involve test takers more directly in the testing process; treat test takers as responsible individuals; provide them with as complete information about the entire testing procedure as possible.

In Ukraine there is a different attitude to language testing in comparison with that typical for western culture. This is due to lack of a test format assessment in both secondary and higher educational systems.

Traditionally assessment of language proficiency has been conducted in the form of examinations based on students' reproduction of

classroom material (e.g., topics, grammar rules) crammed before the exam. There was (and still is) no national testing system to assess language proficiency in simulated real life situations.

Therefore the humanization of the testing process in aviation is important to minimize the test taking ignorance which might negatively influence the test results.

5. Demand accountability for test use; hold yourself, as well as any others who use your test, accountable for the way your test is used.

Accountability is one of the ICAO recommendations concerning rules and procedures to set up an appropriate language testing [4, 5].

6. Recognize that decisions based on test scores are fraught with dilemmas, and that there are no universal answers to these.

This principle can be applied to the ICAO requirement that the total score has to be the lowest among six scores awarded for each of the language profiles presented by the ICAO analytic scale [4].

In practice sometimes it is not easy to assess language performance whether it is at level 3 high and level 4 low. In our research of 124 oral speech samples at level 4 it has been revealed some discrepancy in language performances though all of the samples matched operational level descriptors.

We think it might be due to specificity of test takers' individual cognitive and mnemonic mechanisms responsible for demonstration of fluency, pronunciation, vocabulary range, etc.

Conclusion

Regarding the issues analysed in the article above it can be concluded that the language test for aviation should have highest possible reliability, construct validity and be practical for specific testing situation in aviation. Taking into account that language test in aviation is a high-stake test the test qualities should be considered with respect to specific test tasks and not solely in terms of abstract theories and statistical formulae. It is also important to consider these

qualities from the very beginning of the test planning and development process, and rather than relying solely on ex post facto analyses [3].

Another key issue to ensure quality language testing for aviation is standardisation of the test results. It can be reached by extensive trialing, provision of comparable conditions of test administration, examiner training/auditing based on speech samples, paired/multiple rating, provision of test security and rater/interlocutor qualifications.

References

1. *Convention on International Civil Aviation* (Doc 7300/9). – Annex 1. – ICAO, 2006.
2. *Mell J.* Standards for selection and development of language tests / J. Mell // ICAO-Special Workshop on Language Proficiency, Baku, Azerbaijan, 7th-9th December 2005.
3. *Bachman L.F.* Language Testing in Practice / L.F. Bachman, A.S. Palmer. – Oxford University Press, 1996. – 377 p.
4. *Manual on the Implementation of ICAO Language Proficiency Requirements* / ICAO Doc 9835 AN/453: ICAO Second Edition, 2010. – 186 p.
5. *Language Testing Criteria for Global Harmonization* / ICAO Cir 318 AN/180. – ICAO, 2009. – 38 p.

The editors received the article on 10 February 2011.